DOCUMENT RESUME

ED 090 511 CS 001 064

TITLE	Literacy Among Youths 12-17 Years, United States; Prevalence of Illiteracy as Measured by the Brief Test of Literacy Among Youths by Selected Socioeconomic and Demographic Variables. Vital and Health Statistics Data from the National Health Survey, Series 11, Number 131.
INSTITUTION	National Center for Health Statistics (DHEW), Rockville, Md.
REPORT NO	DHEW-HRA-74-1613
PUB DATE	Dec 73
NOTE	36p.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 ($0.55)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	*Illiteracy; Junior High Schools; Reading; *Reading Ability; Reading Achievement; *Reading Research; Reading Tests; Secondary Education
IDENTIFIERS	*National Health Survey

ABSTRACT

This report presents information on the extent of illiteracy among American youths as assessed by a test specially constructed for this purpose and administered as part of the Health Examination Survey during 1966-70. The test consisted of two parts, one focusing on reading ability and a second one on writing. For the purposes of this report, if a person passed the reading requirement but did not pass the written part he was classified as literate, but if he failed in reading and passed in writing he was termed illiterate. The contents of the report include "Introduction," which discusses sources of data, the psychological test battery, field administration, and quality control; "Findings," which looks at sex, race, age, grade placement, parental education, family income, geographic region, type of community, rate of population change, and number of persons aged 20 or younger in household; "Multivariate Analysis of the Literacy Data," which discusses the development of a method for the analysis of the data; "Consistency of the Literacy Test Results"; "Comparison with other Literacy Studies"; "References"; and "List of Detailed Tables." Appendixes include sample pages from the reading test, statistical notes, and definitions of certain terms used in this report. (WR)

VITAL and HEALTH STATISTICS

DATA FROM THE NATIONAL HEALTH SURVEY

Series 11-Number No. 131

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

NATIONAL CENTER FOR HEALTH STATISTICS
NCHS

Literacy Among Youths 12-17 Years

United States

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
Public Health Service
Health Resources Administration

Series 11 reports present findings from the National Health Examination Survey, which obtains data through direct examination, tests, and measurements of samples of the U.S. population. Reports 1 through 38 relate to the adult program, Cycle I of the Health Examination Survey. The present report is one of a number of reports of findings from the children and youth programs, Cycles II and III of the Health Examination Survey. These latter reports from Cycles II and III are being published in Series 11 but are numbered consecutively beginning with 101. It is hoped this will guide users to the data in which they are interested.

Vital and Health Statistics-Series 11-131

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 - Price 55 cents

Data from the
NATIONAL HEALTH SURVEY

Series 11
Number 131

Literacy Among Youths 12-17 Years

United States

Prevalence of illiteracy as measured by the Brief Test of Literacy among youths by selected socioeconomic and demographic variables.

DHEW Publication No. (HRA) 74-1613

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service

Health Resources Administration
National Center for Health Statistics
Rockville, Md. December 1973

NATIONAL CENTER FOR HEALTH STATISTICS

EDWARD B. PERRIN, Ph.D., *Acting Director*

PHILIP S. LAWRENCE, Sc.D., *Deputy Director*
GAIL F. FISHER, *Assistant Director for Health Statistics Development*
WALT R. SIMMONS, M.A., *Assistant Director for Research and Scientific Development*
JOHN J. HANLON, M.D., *Medical Advisor*
JAMES E. KELLY, D.D.S., *Dental Advisor*
EDWARD E. MINTY, *Executive Officer*
ALICE HAYWOOD, *Information Officer*

DIVISION OF HEALTH EXAMINATION STATISTICS

ARTHUR J. McDOWELL, *Director*
GARRIE J. LOSEE, *Deputy Director*
LINCOLN I. OLIVER, *Chief, Psychological Statistical Branch*
HAROLD J. DUPUY, Ph.D., *Psychological Advisor*
HENRY W. MILLER, *Chief, Operations and Quality Control Branch*

COOPERATION OF THE BUREAU OF THE CENSUS

In accordance with specifications established by the National Health Survey, the Bureau of the Census, under a contractual agreement, participated in the design and selection of the sample, and carried out the first stage of the field interviewing and certain parts of the statistical processing.

Vital and Health Statistics-Series 11-131

DHEW Publication No. (HRA) 74-1613
Library of Congress Catalog Card Number 73-600227

CONTENTS

	Page
Introduction	1
Source of the Data	1
The Psychological Test Battery	2
Field Administration	3
Quality Control	3
Findings	3
Sex	3
Race	4
Age	4
Grade Placement	5
Parental Education	5
Family Income	5
Geographic Region	6
Type of Community	6
Rate of Population Change	7
Number of Persons Aged 20 or Younger in Household	7
Multivariate Analysis of the Literacy Data	7
Consistency of the Literacy Test Results	9
Comparison With Other Literacy Studies	9
References	10
List of Detailed Tables	11
Appendix I. Sample Pages From Reading Test	21
Appendix II. Statistical Notes	22
The Survey Design	22
Reliability	24
Sampling and Measurement Error	24
Small Numbers	25
Imputation	25
The Stepwise Regression Method	25
Appendix III. Definitions of Certain Terms Used in This Report	27

SYMBOLS

Data not available	. . .
Category not applicable	. . .
Quantity zero	-
Quantity more than 0 but less than 0.05	0.0
Figure does not meet standards of reliability or precision	*

LITERACY AMONG YOUTHS 12-17 YEARS

Dorothee K. Vogt, Ph.D., *Division of Health Examination Statistics*

INTRODUCTION

This report presents information on the extent of illiteracy among American youths as assessed by a test specially constructed for this purpose and administered as part of the Health Examination Survey during 1966-70. The decision to include such a test in this particular program of the series of surveys was made because a relationship has long been recognized between certain states of ill health and illiteracy; it was therefore considered desirable to obtain a measure of the prevalence of illiteracy in the population along with relevant medical data. This first report contains the findings on illiteracy rates by sex, age, and certain socioeconomic and demographic variables only. Discussion of the findings in relation to health data will be published in subsequent reports.

A number of attempts have been made in the last few years to define and assess the prevalence of illiteracy in the United States.[8,10] The approaches to the problem vary widely.[12] Sometimes estimated rates of illiteracy are based on household interviews where one household member provides information on the reading or writing ability of other members.[11] In other cases literacy data are related to the ability of a person to perform specific functions, for example, completing application forms.[10] The purpose and unique function of the Health Examination Survey, however, was to use direct examination and therefore a test was developed which could be administered to each sample person. This "Brief Test of Literacy" was specially constructed and pretested for the survey by Thomas F. Donlon and W. Miles McPeek of the Educational Testing Service, Princeton, New Jersey.[2] It adopted a definition of literacy which is commonly used in this country, namely, "literacy is that level of achievement which is attained by the average child in the United States at the beginning of the fourth grade."[1] The items of the test were developed on the basis of this definition and the test was validated on a group of fourth graders. The estimates of illiteracy presented in this report are based on the same definition.

A detailed description of the Brief Test of Literacy and of its development has been published. The test consists of two separate parts, one focusing on reading ability only and a second one on writing. The correlation between scores on the two parts is of course not perfect; this is a reflection of both the nature of the tests and scoring procedures as well as the different mental activities on which they are based. The test developers concluded that the best single indicator of a person's literacy is given by the results of the reading test only. Their recommendation has been accepted for the purposes of this report, i.e., if a person passed the reading requirement but did not pass the written part, he was classified as "literate," whereas if he failed in reading but passed in writing only, he was termed "illiterate." A breakdown of the sample into those who passed or failed one or both tests is given in appendix II.

Source of the Data

The Health Examination Survey is one of the major programs of the National Center for Health Statistics which was authorized under the National Health Survey Act of 1956 by the 84th Congress to be a continuing Public Health Service activity to determine the health status of the

population.[3,4] The National Health Survey is carried out through three different survey programs. One of these is the Health Interview Survey, which is concerned primarily with the impact of illness and disability upon people's lives and actions and the differentials observable in various population groups. It collects information from samples of people by household interviews. A second program, that of the Division of Health Resources Statistics, obtains health data as well as health resource and utilization information through surveys of hospitals, nursing homes, other resident institutions, and the entire range of personnel in health occupations. The third major program is the Health Examination Survey. Literacy test results presented in this report were obtained from the Health Examination Survey of youths completed in 1970.

Data are collected in the Health Examination Survey by direct physical examinations, tests, and measurements performed on selected samples representing specific segments of the U.S. population. This approach provides the best way to obtain actual diagnostic data on the prevalence of certain medically defined illnesses. It is the only way to secure information on unrecognized and undiagnosed conditions and on a variety of physical, physiological, and psychological measures within the population. It also provides demographic and socioeconomic data on the sample population under study.

The Health Examination Survey consists of a series of separate programs referred to as "cycles." Each cycle is limited to some specific segment of the population and to selected aspects of the health of that population. In the first cycle, data were obtained on the prevalence of certain chronic diseases and on the distribution of various measurements and other characteristics in a defined adult population.[3]

The second program, or cycle, involved the selection and examination of a probability sample of the Nation's noninstitutionalized children 6-11 years of age. The examination focused primarily on health factors relating to growth and development, but it also included a psychological test battery and screening for certain diseases and abnormalities. A description of the survey plan, sample design, examination content, and operation of the children's survey is contained in other reports.[4,5] In the third cycle,[6] of which this report is a part, the sampling frame was the same as that used for Cycle II, since this procedure yielded some data on individuals who were examined in both cycles. The target group this time comprised all noninstitutionalized American youths 12-17 years of age, and a probability sample of 7,514 individuals was selected to represent the target population of approximately 22,692,000 youths. Of the sample youths 6,768 were actually examined, a response rate of 90.1 percent. A breakdown of the population sample used in this report into age-sex-race groups is given in appendix I, table I. Survey operations were begun in March 1966 and lasted through March 1970. Each sample person was given a 3-hour examination in the mobile clinic constructed for the survey. It included an examination by a physician and a dentist, tests administered by a psychologist, and a variety of tests, procedures, and measurements performed by technicians.

The Psychological Test Battery

With the exception of the Literacy Test the psychological test battery employed in the survey of youths was the same as in cycle II. This battery is directed towards the assessment of the mental aspects of development and growth. It included the arithmetic computation and reading subtests of the Wide Range Achievement Test to obtain estimates of school achievement in these basic skills. Two subtests of the Wechsler Intelligence Scale for Children were given to yield an approximate measure of the youth's level of intellectual development. The Thematic Apperception Test was chosen for its potential of personality-emotionality measurement, and a modified version of the Goodenough-Harris Drawing Test was administered for additional information on intellectual maturity. As the last one of the 70-minute long series, the Brief Test of Literacy was administered. This test is unique to Cycle III; it really was not applicable to the children 6-11 years old and had not been developed at the time of Cycle I. The reading part of the test, on which the findings in this report are based, consists of seven short passages of 40-50 words each followed by three separate multiple choice questions. These

questions simply reflected a youth's reading comprehension. An example is given in appendix I. The scoring of the questions was done concurrently by the examiner; if the cut-off score, which had been predetermined in the pilot study, was exceeded, the remainder of the test was not given. The cut-off score, using a correction for guessing, was set at 10.5 score points. Thus if in the course of the examination, four passages were correctly answered, this was considered to be sufficient evidence of literacy.

Field Administration

The testing was done during a single visit to the mobile examination center which was designed and set up specially for use in the survey. Prior to the examination, questionnaires were filled out by the youth, his parents, and his school; these provided socioeconomic data, medical and behavioral histories, and certain attitudes. The examination room was well-lighted and temperature-controlled as well as sound-controlled. Two psychologists with at least masters' degrees, previous testing experience, and special preparation for this particular task administered the tests. The fact that the literacy test was the last one of the battery introduced a problem. In some cases examination time had actually run out and it had to be left to the psychologist's judgment to decide whether the examinee was indeed literate or not. Such judgments were rendered for about 200 youths, less than 3 percent of the total sample. A detailed breakdown of the results as coded is contained in appendix II. During the course of the survey, a total of 12 psychologists participated in the administration of the battery.

Quality Control

For the entire psychological test battery, uniformity of testing techniques and conformity with instructions as well as of recording and scoring was maintained at a high level throughout the cycle. Prior to field testing, examiners were thoroughly trained by the Psychological Advisor to the Health Examination Survey in the special standard procedures used in the survey. They were required to memorize the testing instructions. A thorough review of testing practices was also given by the Psychological Advisor during retraining and practice sessions which immediately preceded the start of examination at each stand and on other occasions during the cycle as needed.

The two field psychologists at each examination location exchanged all test forms daily and checked each other's tests for any apparent errors in administration or mistakes in recording.

An entire testing session was taped by each of the field psychologists one day during each week of testing. The transcription of the taped session was reviewed by headquarters staff—the Supervising Field Psychologist of the Psychological Advisor—who noted errors, commented on testing procedures as required, and then returned the tapes with comments to the examiners for study.

All test forms were immediately checked when they arrived at headquarters following the completion of the total round of examinations at any given location.

FINDINGS

Results of the Brief Test of Literacy administered to a national probability sample of youths in the Health Examination Survey of 1966-70 indicate that 4.8 percent of the approximately 23 million noninstitutionalized youths 12-17 years old in the United States are illiterate, i.e., they cannot read at the beginning fourth grade level (table 1). Thus an estimated 1 million teenagers are classified as illiterate under this definition. Distinct differences in rates of illiteracy were found among differing social and economic groups of the population. These differences are discussed below. As in all sample surveys, the estimates obtained are subject to sampling error. The standard errors associated with the rates are given in each of the detailed tables. Sample frequencies for selected variables are given in appendix II.

Sex

The most consistent differences in the literacy rates were found between male and female youths. Table 1 shows the rates of illiteracy by age, race, and sex. The difference between the total rates of 6.7 percent for males and 2.8

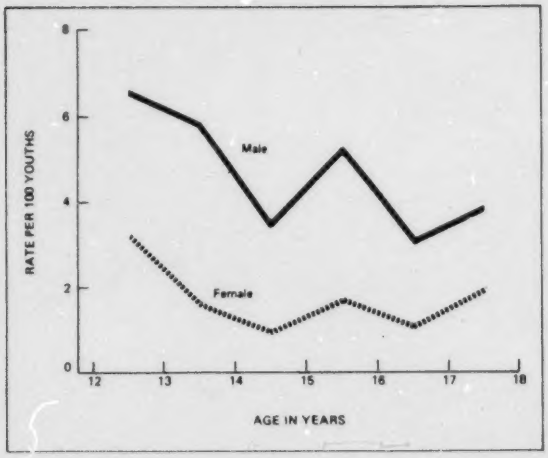


Figure 1. Illiteracy rates for white youths by age and sex, United States, 1966-70.

percent for females is statistically significant, as are the differences for white males (4.7 percent) as compared to white females (1.7 percent) and Negro males (20.5 percent) as compared to Negro females (9.6 percent). If one examines the differences between rates for males and females, they have the same sign at every age level both for white and Negro youths (table 1 and figures 1 and 2). Among white youths there appears to be a leveling off with age, the differences between males and females tend to become smaller. This trend is not maintained for Negro youths particularly in the 17-year-old group. It is not possible to say whether this is due to sampling error or whether some response bias exists in that particular age group. The nature of the test might have caused older teenagers not to take it seriously enough.

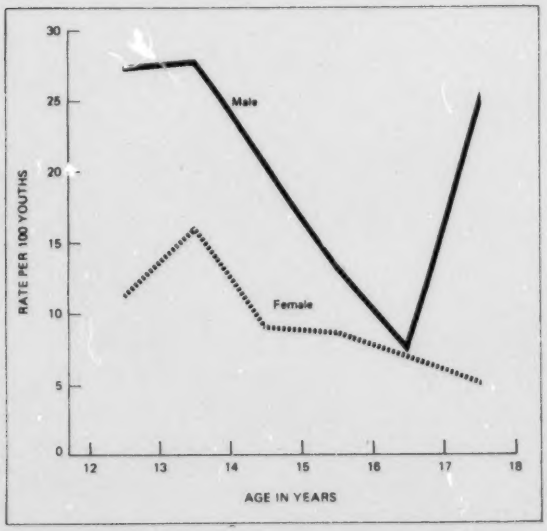


Figure 2. Illiteracy rates for Negro youths by age and sex, United States, 1966-70.

Race

The racial groups in the survey consisted of white, Negro, and "all other" youths. The "all other" group comprised American Indians, Chinese, Japanese, and races other than white or Negro and constituted less than 1 percent of the total sample. No sample youths in that category were found to be illiterate and thus no valid separate estimates of illiteracy for that racial group can be given. The illiteracy rate for Negro youths (15.0 percent) was several times larger than that for white youths (3.2 percent), (table 1 and figure 3).

Age

The survey was designed to have roughly equal numbers of youths in the age groups from 12 to 17 years (appendix table I). Age was defined as age at last birthday. Table 1 shows the rates by age, race, and sex and it appears

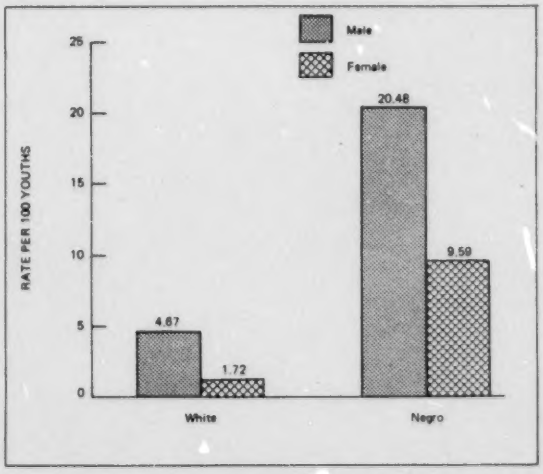


Figure 3. Illiteracy rates by sex and race, United States, 1966-70.

that the rates varied with age, although the test had been constructed to be well below the level of normal 12-year-olds. No single age group shows a significantly higher rate than all others, but there exists a linear trend when the four sex-race groups are considered jointly. However, the fluctuations are large and the improvement with age is by no means consistent, particularly if one looks at the estimates for sex-race groups separately (figures 1 and 2). Because of the inconsistencies and because the trend is only slight, in all further analyses except for grade placement, the age groups have been pooled.

Grade Placement

If 6 years of age is considered as the school starting age, 12-year-olds would be in seventh grade. There are of course variations, but a normally developing child is not usually more than a year above or below that grade at age 12. Thus the bulk of the 6,768 examinees were in grades 6 to 12, and 655 were not in these grades. The estimated rates of illiteracy are given in table 2, and the unweighted sample frequencies are shown in appendix II.

The modal age for seventh graders is 12, for eighth graders 13, and so on. As expected (table 2), the rates of illiteracy for those in or below the modal age groups are small compared to those for youths above modal age. Among those who graduated from high school and went beyond, no illiterates were found. There were also 147 dropouts and 68 youths in special placement for whom the illiteracy rates were high as were the rates for those youths who were below fifth grade, more than 2 years below their "normal" placement.

Parental Education

Among the variables which are generally regarded as good predictors of a child's achievement, parental education ranks high. The survey data also showed a relatively high correlation between test scores on the psychological battery (e.g., the two subtests from the Wechsler Intelligence Scale for Children and two from the Wide Range Achievement Test) and parental education. Rates of illiteracy among youths according to level of education of first-listed parent are given in table 3. The term "first-listed parent" refers to the head of household—usually the father but if not, the adult legally responsible for the youth. The educational groups were combined into broad categories to provide more reliable estimates of illiteracy rates. Sample frequencies on which these rates are based are shown in appendix table II. Estimates are given wherever possible, although in some cases the associated standard errors are very large and the rates have to be interpreted with caution.

In general, higher levels of illiteracy were associated with a lower level of parental education (table 3). This was true for white and Negro youths and for males and females. The rate of illiteracy for those youths whose parents had no formal education, 27.4 percent, is much higher than for those with at least elementary school training, 8.9 percent, and the rate improves more in the other groups (figure 4). It must be borne in mind, of course, that "number of years in school" is not a very good means of comparison between ethnic groups whose total school environment has been radically different and that racial differences persist within educational levels.

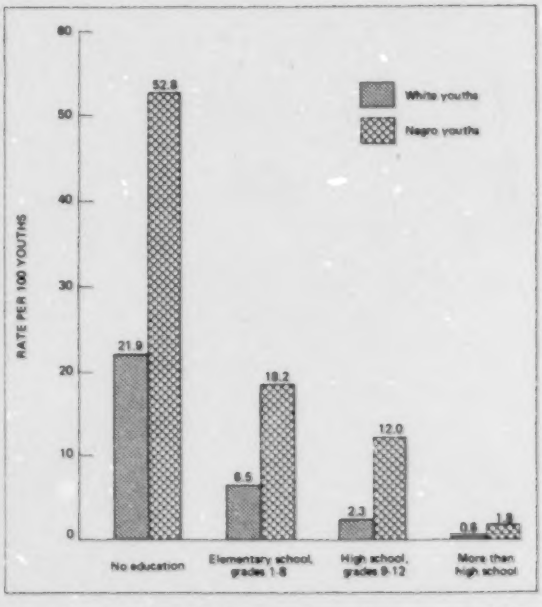


Figure 4. Illiteracy rates by parental education and race, United States, 1966-70.

Family Income

Illiteracy was found to be related to income in the present study. The income groups were based on what was reported as "total combined family income for the preceding 12 months." The rate of illiteracy for each income group is presented in table 4. A downward trend of illiteracy with increasing income is evident for both white and Negro youths. The rate of illiteracy declined from 14 percent in the lowest income group (less than $3,000) to 0.3 percent in the highest ($15,000 or more). Small sample sizes for Negro youths in the higher income groups do not present reliable estimates of illiteracy rates for those groups. The sample frequencies are given in table II. Some relation exists of course between family income and parental education, but the observed correlation in the survey data was only around 0.5 percent. Although parental education is often thought of as the more stable indicator of school achievement (income being more subject to fluctuation), the literacy rates seemed to depend at least as strongly on income levels [see the section on the AID (Automatic Interaction Detector) stepwise regression procedure in appendix II].

Geographic Region

The United States was stratified into four broad geographic regions with approximately equal populations. These regions are defined in appendix III. Table 5 compares illiteracy rates among these regions. It appears that youths living in the South Region have a higher prevalence of illiteracy than youths in other regions (figures 5 and 6). For both races combined the difference between the rate for the South (9.9 percent) and any of the other regions is statistically significant. This is also true for males only (13.6 percent) and for females only (5.9 percent). When the racial groups are examined separately the rates for the South are significantly higher than for the Northeast and Midwest Regions, but because of larger standard errors the differences between the South and West are not statistically significant.

Type of Community

Some effect on literacy could perhaps be attributed to the differences between urban and rural environments. Illiteracy rates for urban and rural places are given in table 6 and figure 7. The combined rate for urban (4.3 percent) was not significantly different from the rural rate (5.6 percent). Looking at the four sex-race groups separately, all the differences point in the same direction though most are not statistically significant—the incidence of illiteracy being lower

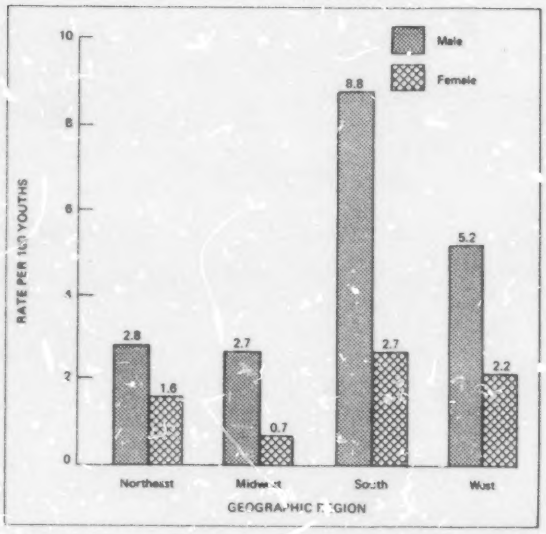


Figure 5. Illiteracy rates for white youths by geographic region, United States, 1966-70.

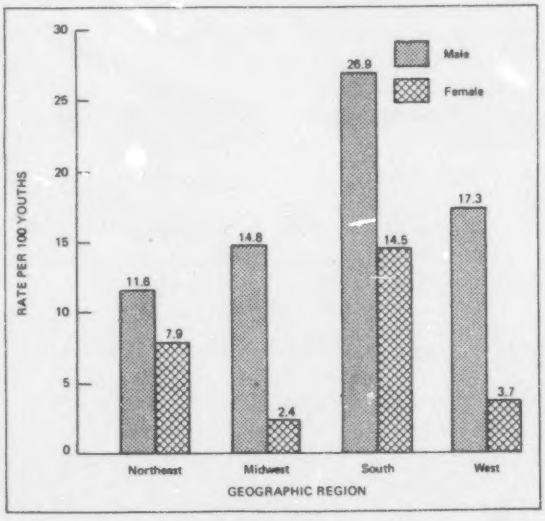


Figure 6. Illiteracy rates for Negro youths by geographic region, United States, 1966-70.

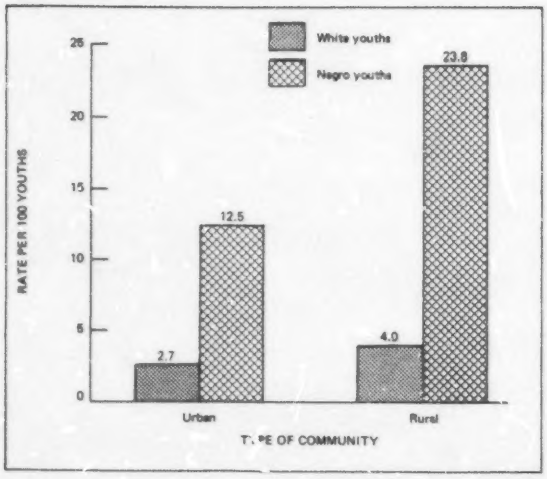


Figure 7. Illiteracy rates for urban and rural communities by race, United States, 1966-70.

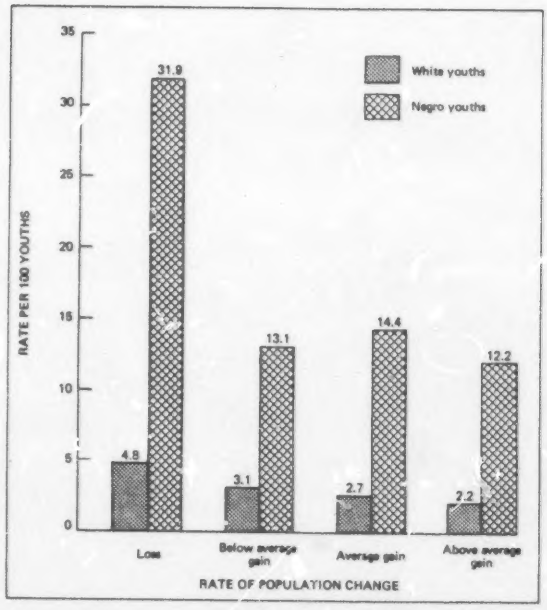


Figure 8. Illiteracy rates by rate of population change and race, United States, 1966-70.

than the national average for each group in the urban areas and higher for each group in the rural ones. For Negro males the difference was significant, however, in spite of the very large error estimates. The towns (urbanized or non-urbanized) can be further broken down by size of city or community as in table 7, but this classification does not show any obvious trends.

Rate of Population Change

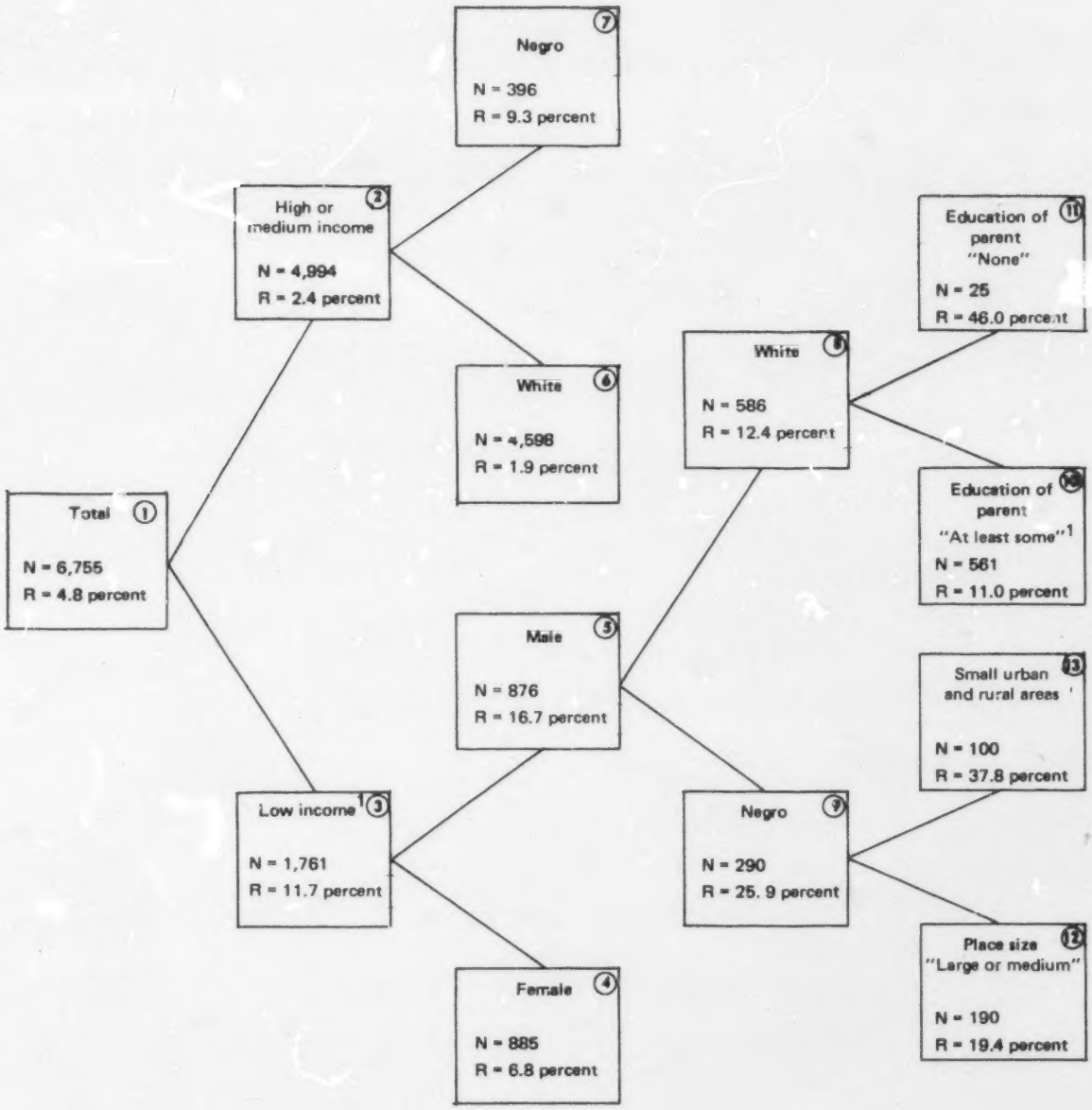
The rate of population change reflects the changes in PSU's between 1950 and 1960. This variable was incorporated into the sample design because of its value as a gross economic indicator, e.g., "above average gain" usually includes new suburban or industrial areas whereas "loss" or "below average gain" is associated with depressed areas, and consequent implications for health data. The rates of illiteracy distributed by the rate of population change are given in table 8 and figure 8. Comparing the four sex-race groups in the extreme categories with their national means does not show statistically significant differences. But again these differences are all in the same direction—slightly higher rates of illiteracy prevail in the category marked "loss" and lower ones in the "above average gain" category.

Number of Persons Aged 20 or Younger in Household

As table 9 shows, a difference exists in the literacy rate between large and small families. Overall a trend can be observed of higher illiteracy rates for larger families although fluctuations exist. When there are six or more youths 20 years of age or younger in a household, the illiteracy rate is significantly higher than that for youths in smaller families. This applies both to males and females but only when the totals for both races are used. When rates are considered for the separate racial groups, differences are still discernible, but the standard errors are too large to conclude that the differences are statistically significant.

MULTIVARIATE ANALYSIS OF THE LITERACY DATA

In an attempt to study literacy as a function of the socioeconomic variables recorded simultaneously in the survey, another approach was used. Most classical methods of multivariate



N = Number of sample youths in group

R = Rate of illiteracy in group

[1] Includes a few cases without response

Figure 9. Result of stepwise regression by means of the AID program.

analysis cannot be applied in the analysis of data such as these since the socioeconomic and demographic variables used in the classifications are in general not quantitative. The Survey Research Center of the University of Michigan Institute for Social Research developed a method for analysis of data which are classified into qualitative categories, the AID (Automatic Interaction Detector) method. The computer program associated with the procedure was modified to take into account the additional complications inherent in the complex sample design used for the Health Examination Survey. Thus it was possible to find out whether much of the illiteracy disclosed in the sample can be directly attributed to a number of socioeconomic factors. The details of this stepwise regression analysis are given in appendix II.

The main conclusions that can be derived are seen in figure 9. They can be summarized as follows: A primary dichotomy was found to exist between income groups, the prevalence of illiteracy being much higher (11.7 percent) at lower income levels than at higher ones (2.4 percent). In the high income group race appeared to contribute more to the variation, while sex accounted for more of the variation in the low income group. Further dichotomizations were seen for white male youths on the basis of parental education and for Negro male youths on the basis of type of residence. However the "explained" variance in terms of these socioeconomic groups constituted only a small fraction of the total since the variation within any one group was large.

CONSISTENCY OF THE LITERACY TEST RESULTS

An estimate of the validity of the literacy test results can be obtained by comparing them with results of other tests in the psychological battery. Analysis indicated that the test most closely associated with literacy was the reading test of the Wide Range Achievement Test (WRAT). Since the literacy test was aimed at the "beginning fourth graders," i.e., should have been passed by the beginning fourth graders, one would expect those who passed this test also to attain at least the fourth grade score on the WRAT reading test. Using the WRAT norm of about 27 raw score points for beginning fourth graders and calling a higher score a "pass" and the rest "fail," the failure rate for this test (about 8 percent) would be higher than the illiteracy rate obtained by the Brief Test of Literacy. However, from the experiences in Cycle II and Cycle III, the WRAT scores obtained in the surveys are somewhat lower than those reported in the original standardization sample. Therefore a score of 25 points rather than 27 might correspond better to the achievement of the average beginning fourth grader. Using this score and again separating the total sample into "pass" and "fail" would result in a failure rate of 5.5 percent, which might lead to the conclusion that the illiteracy rates obtained by the survey are not inconsistent with other data obtained. Nevertheless, further study of the validity of the test is still needed. For this reason and also because any quantitative estimate of literacy and illiteracy generally depends on the definition and measurement used, the results of this report must be interpreted with caution. The findings with respect to socioeconomic and demographic variables may be more stable since similar associations have been established with other types of measurement, e.g., the National Assessment.[8]

COMPARISON WITH OTHER LITERACY STUDIES

Since the inception of the survey, a major program on testing the reading ability of the Nation was carried out by the National Assessment of Educational Progress.[8] The orientation of this project, however, differed substantially from that of the Health Examination Survey (HES) test on literacy. Its scope was much larger both in content and in the number of individuals chosen for the sample. The tests given by National Assessment could be thought of as structured according to a taxonomy such as Bloom's [9] and included cognitive levels of application and evaluation as well as simple knowledge and comprehension. Thus one can only select a fraction of this project as a basis for comparison with the HES results, which are based solely on comprehension. In terms of the Assessment project, the focus of the HES data is on meaning and comprehension. However,

when results of the two studies are compared, good agreement exists in several of the findings:

Female youths have a lower rate of illiteracy than male youths

The South Region shows greater illiteracy than other regions

White youths have lower rates of illiteracy than Negro youths

The rate of literacy improves continuously with increased parental education

As far as size and type of community are concerned, the categories in the two surveys cannot actually be compared. In the HES records the size of an urban community was given but no distinction was made between "inner city" and "suburb." These represent the extreme groups in the National Assessment data, but if they were combined it would seem that the resulting estimate would agree with the one from the HES survey indicating that literacy rates of rural youths are lower than for their urban counterparts.

Other studies of illiteracy have been undertaken in the past few years, but again their focus is too different to warrant any direct comparison, for example, the studies by Louis Harris and Associates[10] define literacy in terms of "survival thresholds." Persons who have low survival thresholds are those who would not survive in a paperwork society. The tests which were developed for this purpose were facsimiles of simplified application forms and instructions. Although the age groups tested in the Harris survey did not coincide with the HES target population, the results exhibited similar trends as far as sex, race, location, and socioeconomic factors are concerned.

The U.S. Bureau of the Census also publishes some estimates of illiteracy in the United States. However, the latest report of their estimates (November 1969) assumed that "everyone who had completed 6 or more years of school was literate."[11] Thus their data cannot be used for direct comparison. Indeed the findings of the HES survey indicate that such definition of literacy might lead to serious underestimates of the number of people in the Nation who cannot be classified as "literate."

REFERENCES

[1] English, H. B., and English, A. C.: *A Comprehensive Dictionary of Psychological and Psychoanalytical Terms.* New York. David McKay Co., 1958.

[2] National Center for Health Statistics: Development of the Brief Test of Literacy. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 27. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1968.

[3] National Center for Health Statistics: Cycle I of the Health Examination Survey, sample and response, United States, 1960-62. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 11-No. 1. Public Health Service. Washington. U.S. Government Printing Office, Apr. 1964.

[4] National Center for Health Statistics: Plan, operation, and response results of a program of children's examinations. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 1-No. 5. Public Health Service. Washington. U.S. Government Printing Office, Oct. 1967.

[5] National Center for Health Statistics: Sample design and estimation procedures for a National Health Examination Survey of children. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 43. Public Health Service. Washington. U.S. Government Printing Office, Aug. 1971.

[6] National Center for Health Statistics: Plan and operation of a Health Examination Survey of U.S. youths 12-17 years of age. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 1-No. 8. Public Health Service. Washington. U.S. Government Printing Office, Sept. 1969.

[7] Guilford, J. P.: *Fundamental Statistics in Psychology,* 4th edition. McGraw-Hill Education Book Co. 1965, p. 382.

[8] National Assessment of Educational Progress: Reading– Summary Report, 02-R-00. May 1972.

[9] Bloom, B. S., ed.: *Taxonomy of Educational Objectives.* New York. David McKay Co., 1956.

[10] National Reading Council: Survival Literacy Studies, conducted by Louis Harris and Associates, 1970-71.

[11] U.S. Bureau of the Census: *U.S. Census of Population, 1970, General Population Characteristics, United States Summary.* Series P-20, No. 217. Washington. U.S. Government Printing Office, Mar. 1971.

[12] Harman, D.: Illiteracy, an Overview. *Harvard Educational Review.* 40(2): 226-243. Cambridge, Mass., May 1970.

○○○

LIST OF DETAILED TABLES

		Page
Table 1.	Estimated illiteracy rates and associated standard errors for youths by age, race, and sex: United States, 1966-70	12
2.	Estimated rates of illiteracy and associated standard errors for youths by age and grade placement in school: United States, 1966-70	13
3.	Estimated illiteracy rates and associated standard errors for youths 12-17 years by education of first-listed parent, race, and sex: United States, 1966-70	14
4.	Estimated illiteracy rates and associated standard errors for youths 12-17 years by family income, race, and sex: United States, 1966-70	15
5.	Estimated illiteracy rates and associated standard errors for youths 12-17 years by geographic region, race, and sex: United States, 1966-70	16
6.	Estimated illiteracy rates and associated standard errors for youths 12-17 years by type of community, race, and sex: United States, 1966-70	17
7.	Estimated illiteracy rates and associated standard errors for youths 12-17 years by type and size of place of residence, race, and sex: United States, 1966-70	18
8.	Estimated illiteracy rates and associated standard errors for youths 12-17 years by rate of population change, race, and sex: United States, 1966-70	19
9.	Estimated illiteracy rates and associated standard errors for youths 12-17 years by number of persons age 20 or younger in household, race, and sex: United States, 1966-70	20

Table 1. Estimated illiteracy rates and associated standard errors for youths by age, race, and sex: United States, 1966-70

Race and sex	Total	Age in years					
		12	13	14	15	16	17
	Rate per 100 youths						
All races							
Both sexes	4.8	6.9	6.2	3.8	4.4	2.7	4.4
Male	6.7	9.4	8.7	5.5	6.2	3.5	6.5
Female	2.8	4.3	3.6	2.0	2.6	1.8	2.3
White							
Both sexes	3.2	4.9	3.8	2.2	3.5	2.0	2.8
Male	4.7	6.6	5.8	3.4	5.2	3.0	3.8
Female	1.7	3.2	1.6	0.9	1.7	1.0	1.9
Negro							
Both sexes	15.0	19.3	21.7	14.5	11.0	7.0	14.8
Male	20.5	27.2	27.7	20.5	13.3	7.3	25.0
Female	9.6	11.2	16.0	8.9	8.6	6.8	5.0
	Standard error						
All races							
Both sexes	0.40	0.71	0.84	0.56	0.64	0.53	0.73
Male	0.57	1.11	1.36	0.92	0.78	0.72	1.33
Female	0.39	0.77	0.81	0.51	1.06	0.62	0.68
White							
Both sexes	0.38	0.85	0.76	0.64	0.60	0.44	0.65
Male	0.58	1.16	1.47	1.00	0.90	0.53	1.09
Female	0.29	0.95	0.31	0.49	0.84	0.58	0.76
Negro							
Both sexes	2.24	2.42	4.30	1.47	3.43	2.82	4.32
Male	2.66	4.15	5.52	2.52	4.14	3.61	7.52
Female	2.46	2.93	5.33	1.90	4.70	2.83	3.07

Table 2. Estimated rates of illiteracy and associated standard errors for youths by age and grade placement in school: United States, 1966-70

Age	Total	Grade placement										Drop-out	Special placement
		None-fourth	Fifth	Sixth	Seventh	Eighth	Ninth	10th	11th	12th	Beyond 12th		
	Rate per 100 youths												
Total	4.8	80.2	43.3	11.7	5.9	3.2	2.4	1.3	0.6	0.7	-	20.5	74.8
12 years	6.9	*	37.7	7.1	2.8	0.8	*	...	...	...	...	...	81.2
13 years	6.2	*	*	28.3	6.6	1.5	-	*	*	...	...	...	*
14 years	3.8	*	*	*	14.2	3.2	0.8	-	*	*	...	*	*
15 years	4.4	*	*	*	27.7	22.5	4.7	0.3	0.7	*	*	*	*
16 years	2.7	*	*	...	*	*	3.3	1.8	0.1	0.6	*	21.0	*
17 years	4.4	...	...	...	*	*	*	9.7	1.5	0.7	-	20.0	*
	Standard error												
Total	0.40	16.65	7.81	1.99	0.83	0.41	0.89	0.26	0.20	0.37	-	3.40	5.33
12 years	0.71	24.24	8.98	1.85	0.63	0.80	---	---	---	---	---	---	10.59
13 years	0.84	---	---	4.76	2.02	0.33	-	---	---	---	---	---	---
14 years	0.56	---	---	---	3.88	1.07	0.43	-	---	---	---	---	---
15 years	0.64	---	---	---	14.96	8.09	1.48	0.18	0.88	---	---	---	---
16 years	0.53	---	---	---	---	---	2.22	0.66	0.11	0.58	---	4.63	---
17 years	0.73	---	---	---	---	---	---	3.32	0.68	0.45	---	4.19	---

Table 3. Estimated illiteracy rates and associated standard errors for youths 12-17 years by education of first-listed parent, race, and sex: United States, 1966-70

Race and sex	Total	Education of first-listed parent				
		No formal education	Elementary school	High school	College	Unknown
	Rate per 100 youths					
All races						
Both sexes	4.8	27.4	8.9	3.5	0.6	9.2
Male	6.7	36.4	12.9	4.9	0.8	11.5
Female	2.8	16.2	5.2	1.9	0.4	7.1
White						
Both sexes	3.2	21.9	6.5	2.3	0.6	6.3
Male	4.7	33.4	9.6	3.4	0.7	8.4
Female	1.7	8.6	3.7	1.1	0.4	3.8
Negro						
Both sexes	15.0	52.8	18.2	12.0	1.8	16.1
Male	20.5	47.8	25.1	16.9	3.4	22.7
Female	9.6	62.0	11.2	7.3	-	12.6
	Standard error					
All races						
Both sexes	0.40	11.18	0.97	0.37	0.20	2.29
Male	0.57	19.15	1.44	0.60	0.33	3.72
Female	0.39	14.23	0.81	0.37	0.20	2.87
White						
Both sexes	0.38	12.19	0.76	0.38	0.19	2.34
Male	0.58	23.87	1.23	0.65	0.33	3.31
Female	0.29	12.75	0.65	0.21	0.21	3.81
Negro						
Both sexes	2.24	21.37	3.35	1.45	2.36	5.53
Male	2.66	30.11	3.80	1.93	4.98	11.18
Female	2.46	35.65	3.47	2.68	-	5.06

Table 4. Estimated illiteracy rates and associated standard errors for youths 12-17 years by family income, race, and sex: United States, 1966-70

Race and sex	Total	Family income						
		Less than $3,000	$3,000-$4,999	$5,000-$6,999	$7,000-$9,999	$10,000-$14,999	$15,000 or more	Unknown
All races	Rate per 100 youths							
Both sexes	4.8	14.3	9.3	4.8	1.9	0.9	0.3	6.6
Male	6.7	19.9	14.1	6.1	2.3	1.7	0.4	9.5
Female	2.8	9.4	4.4	3.5	0.8	0.2	0.3	3.1
White								
Both sexes	3.2	9.8	7.4	3.5	1.9	0.8	0.4	5.0
Male	4.7	14.1	11.5	4.8	2.8	1.3	0.4	7.3
Female	1.7	6.3	3.3	2.3	0.7	0.2	0.3	2.1
Negro								
Both sexes	15.0	22.1	14.9	12.6	3.1	4.7	*	16.6
Male	20.5	29.0	22.2	14.7	3.9	11.4	*	25.7
Female	9.6	15.3	7.5	10.6	2.1	-	*	8.6
All races	Standard error							
Both sexes	0.40	2.14	1.22	0.73	0.37	0.25	0.19	2.00
Male	0.57	2.58	1.73	1.09	0.76	0.51	0.26	2.83
Female	0.39	2.15	1.07	0.82	0.33	0.05	0.31	1.62
White								
Both sexes	0.38	1.73	1.21	0.80	0.36	0.19	0.20	1.84
Male	0.58	2.86	1.52	1.29	0.80	0.39	0.26	2.58
Female	0.29	1.74	1.02	0.66	0.34	0.06	0.31	1.60
Negro								
Both sexes	2.24	4.07	2.99	1.79	2.14	2.92	*	7.67
Male	2.66	3.48	4.63	2.62	2.85	6.36	*	11.76
Female	2.46	5.45	2.49	4.07	2.12	-	*	5.62

Table 5. Estimated illiteracy rates and associated standard errors for youths 12-17 years by geographic region, race, and sex: United States, 1966-70

Race and sex	United States	Geographic region			
		Northeast	Midwest	South	West
All races	Rate per 100 youths				
Both sexes	4.8	3.2	2.3	9.9	4.2
Male	6.7	3.8	3.8	13.6	6.1
Female	2.8	2.5	0.9	5.9	2.3
White					
Both sexes	3.2	2.2	1.7	5.9	3.7
Male	4.7	2.8	2.7	8.8	5.2
Female	1.7	1.6	0.7	2.7	2.2
Negro					
Both sexes	15.0	9.5	9.0	20.7	10.9
Male	20.5	11.6	14.8	26.9	17.3
Female	9.6	7.9	2.4	14.5	3.7
All races	Standard error				
Both sexes	0.40	0.73	0.44	1.35	1.08
Male	0.57	1.15	0.87	1.73	1.64
Female	0.39	0.33	0.43	1.41	0.91
White					
Both sexes	0.38	0.62	0.43	0.78	1.11
Male	0.58	0.96	0.75	1.24	1.55
Female	0.29	0.34	0.40	0.44	1.01
Negro					
Both sexes	2.24	2.03	2.02	4.07	3.86
Male	2.66	4.30	2.86	4.09	8.21
Female	2.46	1.76	1.72	5.06	4.79

Table 6. Estimated illiteracy rates and associated standard errors for youths 12-17 years by type of community, race, and sex: United States, 1966-70

Race and sex	Total	Type of community	
		Urban	Rural
	Rate per 100 youths		
All races			
Both sexes	4.8	4.3	5.6
Male	6.7	6.0	7.8
Female	2.8	2.5	3.2
White			
Both sexes	3.2	2.7	4.0
Male	4.7	4.0	5.7
Female	1.7	1.4	2.3
Negro			
Both sexes	15.0	12.5	23.8
Male	20.5	16.9	32.7
Female	9.6	8.3	14.5
	Standard error		
All races			
Both sexes	0.40	0.34	0.83
Male	0.57	0.63	0.89
Female	0.39	0.21	1.04
White			
Both sexes	0.38	0.40	0.69
Male	0.58	0.68	0.86
Female	0.29	0.24	0.67
Negro			
Both sexes	2.24	1.26	7.96
Male	2.66	2.21	6.80
Female	2.46	1.27	10.69

Table 7. Estimated illiteracy rates and associated standard errors for youths 12-17 years by type and size of place of residence, race, and sex: United States, 1966-70

Race and sex	Total	Urbanized area				Nonurbanized area			
		3 million or more	1-2.9 million	250,000-999,999	Less than 250,000	25,000 or more	10,000-24,999	2,500-9,999	Rural
	Rate per 100 youths								
All races									
Both sexes---	4.8	4.3	2.1	5.1	6.7	2.9	6.6	4.6	5.6
Male---	6.7	5.7	3.2	6.7	10.7	5.5	7.2	6.9	7.8
Female---	2.8	2.9	1.0	3.4	2.8	0.5	6.0	2.7	3.2
White									
Both sexes---	3.2	3.0	1.1	2.9	3.2	3.0	4.4	3.3	4.0
Male---	4.7	4.0	1.5	4.9	5.6	5.7	4.9	4.4	5.7
Female---	1.7	2.0	0.6	0.9	1.0	0.5	3.9	2.3	2.3
Negro									
Both sexes---	15.0	8.5	9.1	17.2	18.6	*	31.0	17.2	23.8
Male---	20.5	12.1	13.3	18.2	24.9	*	34.0	33.0	32.7
Female---	9.6	5.5	4.0	16.3	10.1	*	28.0	5.9	14.5
	Standard error								
All races									
Both sexes---	0.40	0.32	0.59	2.20	4.56	1.84	4.97	0.82	0.83
Male---	0.57	0.62	0.99	2.74	7.46	3.41	6.61	1.96	0.89
Female---	0.39	0.27	0.34	1.67	1.85	0.54	3.93	0.75	1.04
White									
Both sexes---	0.38	0.56	0.46	0.93	1.88	1.90	5.03	1.29	0.69
Male---	0.58	0.88	0.54	1.77	3.44	3.53	5.88	1.96	0.86
Female---	0.29	0.40	0.44	0.50	1.04	0.57	4.55	1.29	0.67
Negro									
Both sexes---	2.24	1.14	2.84	3.28	4.94	---	17.77	4.84	7.96
Male---	2.66	1.92	4.41	7.47	9.03	---	25.01	5.16	6.80
Female---	2.46	1.53	3.98	5.12	1.81	---	32.62	2.60	10.69

Table 8. Estimated illiteracy rates and associated standard errors for youths 12-17 years by rate of population change, race, and sex: United States, 1966-70

Race and sex	Total	Rate of population change			
		Loss	Below average gain	Average gain	Above average gain
	Rate per 100 youths				
All races					
Both sexes	4.8	6.0	4.4	5.6	3.3
Male	6.7	7.6	6.3	8.6	4.8
Female	2.8	4.5	2.4	2.7	1.5
White					
Both sexes	3.2	4.8	3.1	2.7	2.2
Male	4.7	6.7	4.8	3.7	3.5
Female	1.7	3.0	1.3	1.6	0.8
Negro					
Both sexes	15.0	31.9	13.1	14.4	12.2
Male	20.5	30.9	15.4	24.2	16.6
Female	9.6	32.6	10.4	5.8	7.4
	Standard error				
All races					
Both sexes	0.40	1.42	1.34	1.07	0.72
Male	0.57	1.59	1.89	1.91	0.99
Female	0.39	1.51	0.88	0.46	0.52
White					
Both sexes	0.38	0.83	1.22	0.46	0.52
Male	0.58	1.26	1.88	0.82	0.80
Female	0.29	0.81	0.75	0.31	0.31
Negro					
Both sexes	2.24	21.96	3.29	2.17	2.65
Male	2.66	20.54	3.30	4.01	2.42
Female	2.46	22.99	4.55	0.96	3.93

Table 9. Estimated illiteracy rates and associated standard errors for youths 12-17 years by number of persons age 20 or younger in household, race, and sex: United States, 1966-70

Race and sex	Total	Number of persons age 20 or younger in household					
		1	2	3	4	5	6 or more
All races	Rate per 100 youths						
Both sexes	4.8	3.0	2.8	3.7	4.4	5.0	10.2
Male	6.7	4.5	3.8	5.2	6.5	8.2	13.6
Female	2.8	1.5	1.7	2.2	2.4	1.6	6.8
White							
Both sexes	3.2	2.4	1.9	2.7	3.1	3.7	6.9
Male	4.7	4.0	2.5	3.9	4.6	6.4	9.2
Female	1.7	0.8	1.1	1.3	1.6	1.0	4.7
Negro							
Both sexes	15.0	8.5	14.2	15.8	13.4	11.1	18.7
Male	20.5	10.1	19.9	21.8	18.2	16.9	25.1
Female	9.6	7.2	8.5	11.0	7.8	4.6	12.4
All races	Standard error						
Both sexes	0.40	0.58	0.51	0.34	0.73	0.95	1.46
Male	0.57	1.29	0.88	0.73	1.05	1.72	1.69
Female	0.39	0.55	0.41	0.57	0.69	0.66	1.96
White							
Both sexes	0.38	0.56	0.42	0.39	0.80	1.08	1.27
Male	0.58	1.24	0.65	0.63	1.12	2.16	1.44
Female	0.29	0.45	0.36	0.47	0.76	0.59	1.91
Negro							
Both sexes	2.24	3.38	3.72	4.36	2.50	3.04	4.29
Male	2.66	6.12	7.58	8.30	4.40	4.24	4.42
Female	2.46	4.10	3.13	5.65	2.73	2.77	4.97

APPENDIX I
SAMPLE PAGES FROM READING TEST

It was a beautiful gift, wrapped with bright red paper and tied with silver string. It was small, but very heavy. No one knew who had brought it, but it had Mr. Jones' name on top. Mr. Jones just smiled and said, "I'll open it when I get home."

01. Whose name was on the top of the gift?

(A) Mr. Jones
(B) Mr. Pike
(C) Willy
(D) The postman
(E) No one knew

02. In what color paper was the gift wrapped?

(A) Red
(B) Silver
(C) Green
(D) Orange
(E) Yellow

03. Where was the gift going to be opened?

(A) Where it was found
(B) At the police station
(C) In the car
(D) At the office
(E) At home

DO NOT TURN THE PAGE
UNTIL YOU ARE TOLD
TO DO SO.

-0-

It was spring. The young boy breathed the warm air, threw off his shoes, and began to run. His arms swung. His feet hit sharply and evenly against the ground. At last he felt free.

1. What time of the year was it?

(A) Summer
(B) Fall
(C) Spring
(D) December
(E) July

2. What was the young boy doing?

(A) Running
(B) Jumping
(C) Going to sleep
(D) Driving a car
(E) Fighting

3. How did he feel?

(A) Hot
(B) Free
(C) Angry
(D) Cold
(E) Unhappy

DO NOT TURN THE PAGE
UNTIL YOU ARE TOLD
TO DO SO.

-1-

There were footsteps and a knock at the door. Everyone inside stood up quickly. The only sound was that of the pot boiling on the stove. There was another knock. No one moved. The footsteps on the other side of the door could be heard moving away.

4. The people inside the room

(A) Hid behind the stove
(B) Stood up quickly
(C) Ran to the door
(D) Laughed out loud
(E) Began to cry

5. What was the only sound in the room?

(A) People talking
(B) Birds singing
(C) A pot boiling
(D) A dog barking
(E) A man shouting

6. The person who knocked at the door finally

(A) Walked into the room
(B) Sat down outside the door
(C) Shouted for help
(D) Walked away
(E) Broke down the door

DO NOT TURN THE PAGE
UNTIL YOU ARE TOLD
TO DO SO.

-2-

Helen liked going to the movies. Sometimes she went four times a week. Everyone said she was crazy. Why did she always want to go out and spend money, they said, when she could stay home and watch television?

7. What did Helen like to do?

(A) She liked to eat
(B) She liked to swim
(C) She liked to watch baseball
(D) She liked to watch movies
(E) She liked to watch wrestling matches

8. What did people think about her?

(A) They thought she was crazy
(B) They thought she was very smart
(C) They thought she was very nice
(D) They thought she was ugly
(E) They thought she was very old

9. What did people think she should do?

(A) Write a book
(B) Watch television
(C) Go on a diet
(D) Dye her hair
(E) Stop talking so much

DO NOT TURN THE PAGE
UNTIL YOU ARE TOLD
TO DO SO.

-3-

APPENDIX II

STATISTICAL NOTES

The Survey Design

The sample design for the third cycle of the Health Examination Survey, similar to those used for the previous cycles, was that of a multistage, stratified probability sample of loose clusters of persons in land-based segments. Successive elements dealt with in the process of sampling are the primary sampling unit (PSU), census enumeration district (ED), segment, household, eligible youth (EY), and sample youth (SY).

At the first stage, the nearly 2,000 PSU's into which the United States (including Hawaii and Alaska) had been divided and then grouped into 357 strata for use in the Current Population Survey and Health Interview Survey were further grouped into 40 superstrata for use in Cycle III of the Health Examination Survey. The average size of each Cycle III stratum was 4.5 million persons, and all fell between the limits of 3.5 and 5.5 million. Grouping into 40 strata was done in a way that maximized homogeneity of the PSU's included in each stratum, particularly with regard to the degree of urbanization, geographic proximity, and degree of industrialization. The 40 strata were classified into four broad geographic regions (each with 10 strata) of approximately equal population and cross-classified into four broad population density groups (each having 10 strata). Each of the 16 cells contained either two or three strata. A single stratum might include only one PSU, only part of a PSU (for example, New York City, which represented two strata), or several score PSU's.

To take account of the possible effect that the rate of population change between 1950 and 1960 censuses might have had on health, the 10 strata within each region were further classified into four classes ranging from those with population loss to those with the greatest relative increase. Each such class contained two or three strata.

One PSU was then selected from each of the 40 strata. A controlled selection technique was used in which the probability of selection of a particular PSU was proportional to its 1960 population. In the controlled selection an attempt was also made to maximize the spread of the PSU's among the States. While not every one of the 64 cells in the 4x4x4 grid contributes a PSU to the sample of 40 PSU's, the controlled selection technique ensured the sample's matching the marginal distributions in all three dimensions and being closely representative of all cross-classifications.

Generally, within a particular PSU, 20 ED's were selected with the probability of selection of a particular ED proportional to its population in the age groups 5-9 years in the 1960 census, which by 1966 roughly approximated the population in the target age group for Cycle III. A similar method was used for selecting one segment (clusters of households) in each ED. Each of the resultant 20 segments was either a bounded area or a cluster of households (or addresses). All of the youths in the age range properly resident at the address visited were EY's. Operational considerations made it necessary to reduce the number of prospective examinees at any one location to a maximum of 200. The EY's to be excluded for this reason from the SY group were determined by systematic subsampling.

Tables I-III show the number of sample youths used for this study. Table I gives a breakdown into age-race-sex groups, table II shows the frequencies by socioeconomic variables, and table III gives a distribution by age and grade in school.

Reliability

Measurement processes employed in the survey were highly standardized and closely controlled. Of course, this does not mean that the correspondence between the real world and the survey results is exact. Data from the survey are imperfect for three major reasons: Results are subject to sampling error, the actual conduct of a survey never agrees perfectly with the design, and the measurement processes themselves are inexact even though standardized and controlled.

Data recorded for each sample youth are inflated in the estimation process to characterize the larger universe of which the sample youth is representative. The weights used in this inflation process are a product of the reciprocal of the probability of selecting the youth, an adjustment for nonresponse cases, and a poststratified ratio adjustment which increases precision by bringing survey results into closer alignment with known United States population figures by race and sex within single years of age 12-17.

In the third cycle of the Health Examination Survey the sample was the result of three stages of selection—the single PSU from each stratum, the 20 segments from each sample PSU, and the sample youths from the eligible youths. The probability of selecting an individual

Table I. Sample frequency distribution of youths by age, race, and sex: Health Examination Survey, 1966-70

Race and sex	Age in years						
	Total	12	13	14	15	16	17
	Number						
All races							
Both sexes	6,768	1,190	1,208	1,204	1,116	1,092	958
Male	3,545	643	626	618	613	556	489
Female	3,223	547	582	586	503	536	469
White							
Both sexes	5,735	995	1,032	1,011	950	937	810
Male	3,047	540	542	527	525	496	417
Female	2,688	455	490	484	425	441	393
Negro							
Both sexes	999	189	171	189	157	150	143
Male	479	101	80	88	84	57	69
Female	520	88	91	101	73	93	74
All other							
Both sexes	34	6	5	4	9	5	5
Male	19	2	4	3	4	3	3
Female	15	4	1	1	5	2	2

youth is the product of the probabilities of selection at each stage.

Since the strata are roughly equal in population size and a nearly equal number of sample youths were examined in each of the sample PSU's, the sample design is essentially self-weighting with respect to the target population, that is, each youth 12-17 years had about the same probability of being drawn into the sample.

The adjustment upward for nonresponse is intended to minimize the impact of this factor on final estimates by imputing to nonrespondents the characteristics of similar respondents. Here similar respondents were judged to be examined youths in a sample PSU having the same age (in years) and sex as youths not examined in that sample PSU.

The poststratified ratio adjustment used in the third cycle achieved most of the gains in precision which would have been attained if the sample had been drawn from a population stratified by age, race, and sex and made the final sample estimates of population agree exactly with independent controls prepared by the U.S. Bureau of the Census for the noninstitutional population of the United States as of March 9, 1968 (approximate midsurvey point) by race and sex for each single year of age 12-17. The weight of every responding sample youth in each of the 24 age, race, and sex classes is adjusted upward or downward so that the weighted total within the class equals the independent population control for each survey.

Sampling and Measurement Error

The probability design of the survey makes possible the calculation of sampling errors. The sampling error is used here to determine how imprecise the survey test results may be because they come from a sample rather than from the measurement of all elements in the universe.

The estimation of sampling errors for a study of the type of the Health Examination Survey is difficult for at least three reasons: (1) measurement error and "pure" sampling error are confounded in the data—it is not easy to find a procedure which will either completely include both or treat one or the other separately, (2) the survey design and estimation procedure are complex and accordingly require computationally involved techniques for the calculation of variances, and (3) thousands of statistics are computed from the survey, many for subclasses of the population for which there are a small number of cases. Estimates of sampling error are obtained from the sample data and are themselves subject to sampling error which may be large when the number of cases in a cell is small or even occasionally when the number of cases is substantial.

Table II. Unweighted sample frequencies of selected socioeconomic and demographic variables by race: Health Examination Survey, 1966-70

Variable	White	Negro	All other
Total	5,735	999	34
Education of first-listed parent			
No formal education	70	18	-
Elementary school	1,370	429	10
High school	2,885	437	10
College	1,279	54	14
Unknown	131	61	-
Family income			
Less than $3,000	491	326	-
$3,000-4,999	665	275	6
$5,000-6,999	922	158	5
$7,000-9,999	1,440	105	10
$10,000-14,999	1,221	51	5
$15,000 or more	629	15	8
Unknown	367	69	-
Geographic region			
Northeast	1,428	210	3
Midwest	1,582	166	-
South	1,206	498	9
West	1,519	125	22
Type of community			
Urban	3,680	773	23
Rural	2,055	226	11
Rate of population change			
Loss	1,512	80	9
Below average gain	1,463	281	10
Average gain	1,317	454	2
Above average gain	1,443	184	13
Number of youths age 20 or younger in household			
1	774	86	1
2	1,337	125	12
3	1,351	129	7
4	939	170	6
5	550	138	2
6 or more	784	351	6
Grade placement			
Dropout	111	36	-
Special placement	41	27	-
None-fourth grade	18	7	-
Fifth grade	54	20	-
Sixth grade	313	76	1
Seventh grade	875	175	8
Eighth grade	1,083	174	4
Ninth grade	937	171	5
10th grade	920	145	7
11th grade	770	102	3
12th grade	513	59	5
Beyond 12th grade	100	7	1

Table III. Unweighted sample frequencies by grade placement in school and age: Health Examination Survey, 1966-70

Age	Total	Grade placement										Drop-out	Special placement
		None-fourth	Fifth	Sixth	Sev-enth	Eighth	Ninth	10th	11th	12th	Be-yond 12th		
	Number												
Total	6,768	25	74	390	1,058	1,261	1,113	1,072	875	577	108	147	68
12 years	1,190	11	51	296	666	143	2	-	-	-	-	-	21
13 years	1,208	6	16	74	273	697	122	3	1	-	-	-	16
14 years	1,204	6	1	16	84	332	599	144	2	1	-	5	14
15 years	1,116	1	4	4	22	70	296	576	117	3	2	12	9
16 years	1,092	1	2	-	8	14	76	279	539	120	4	46	3
17 years	958	-	-	-	5	5	18	70	216	453	102	84	5

Estimates of approximate sampling variability for selected statistics have been prepared by a replication technique which yields overall variability through observation of variability among random subsamples of the total sample. This method reflects both "pure" sampling variance and a part of the measurement variance. These standard errors of estimates are shown in the detailed tables.

In accordance with usual practice, the interval estimate for the rates given in this report may be considered the range within one standard error of the tabulated statistic with 68-percent confidence or the range within two standard errors of the tabulated statistic with 95-percent confidence. The latter is used as the level of significance in this report.

An approximation of the standard error of a difference $d = x - y$ of two statistics x and y is given by the formula $S_d = (S_x^2 + S_y^2)^{1/2}$ where S_x and S_y are the sampling errors, respectively, of x and y. When the two measures are correlated, this formula will give an overestimate or underestimate of the actual standard error.

Small Numbers

In some tables statistics are shown for cells for which the sample size is so small that the sampling error may be several times as great as the statistic itself. Such numbers, if shown, have been included in the belief that they may help to convey an impression of the overall story of the table.

Imputation

In addition to those youths who were selected in the sample but not examined, there were some whose examination was incomplete in one procedure or another. The effect of nonresponse on HES data is particularly serious, since each SY is weighted in proportion to his representation in the Nation. To minimize the need for reweighting every part of the long examination, other available information on SY's with missing items was used to obtain a substitute or imputed response.

In the case of the literacy test the information used for missing data was based on the test scores most closely associated with the concept of literacy, i.e., the reading test of the Wide Range Achievement Test and, to a lesser degree, the vocabulary test of the Wechsler Intelligence Scale for Children. If one examines the distribution of WRAT sample scores for literate youths and illiterate youths separately, one finds that they differ as to means and range. Although they are not strictly normally distributed, they are unimodal and the method of "equal likelihood" outlined in Guilford[7] was used to obtain a "cut-off point" such that a person with a score below it would be termed illiterate, above it as literate, and with that cut-off score itself as having a 50/50 chance of being in either category. This method seemed to yield satisfactory results since the ratio of illiterate classifications to the literate ones in the imputed group turned out to be comparable to that based on examination results. The actual sample numbers as obtained were

	Number	*Percent*
Total	6,768	100.0
Literate (by test)	6,151	91.8
Illiterate(by test)	263	3.6
Literate (examiner's opinion)	100	1.4
Illiterate (examiner's opinion)	17	0.2
Too illiterate to take test	51	0.8
Physical disabilities	12	0.2
No response available	174	2.0

For purposes of analysis all literate categories were grouped together as were the categories for illiteracy. The records of the physically handicapped were reexamined and most were classified as literate or illiterate. For the 174 youths for whom no data were available imputations were made except in a very few atypical cases. Thus the final sample breakdown became

	Number
Total	6,768
Literate	6,413
Illiterate	342
Physical disabilities	5
No imputation possible	8

The Stepwise Regression Method

A computer program called AID (Automatic Interaction Detector) was written at the Survey Research Center of the University of Michigan, and it was adopted for the National Center for Health Statistics to take account of the weighting techniques applied to the data obtained by the survey. With this program categorical data can be analyzed, and it was used to examine the literacy results.

The assumption made is that literacy Y is a function of "predictor variables" $X_1, X_2 \ldots\ldots X_n$ which may or may not be correlated among themselves. A dichotomization or "split" is performed on that predictor variable X_i which accounts for the greatest proportion of the total sum of squares of Y. After this split has been accomplished, the resulting groups are dichotomized in a similar manner, i.e., the variable which accounts for most of the variation is selected in each group separately and the groups are split accordingly. This process is continued until one of several specified restrictions come into force. One restriction is the minimum group size which prevents a splitting into groups which are too small to be reliable. The parameter $P1$ limits the process to groups for which the total sum of squares is greater than a fixed proportion $P1$ of the overall sum of squares and $P2$ terminates the dichotomization if the "between sum of squares" resulting from it is less than a proportion $P2$ of the overall sum. Figure 9 in the text illustrates the result of the analysis on the literacy data Y based on the assumption that literacy is dependent on the following predictor variables: age, sex, race, family income, parent's education, size of place of residence and whether in an SMSA or not, geographic region, rate of population change, and number of people aged 20 or under in household. In this analysis the minimum group size of 20 and the parameter $P1 = .01$, which specified that a group with less than 1 percent of the total sum of squares should not be split any further, did not stop the process. In each terminal group, however, the process was stopped by the parameter $P2 = .005$, which restricted further dichotomization to groups for which the "between sum of squares" or the amount of explained variability was less than one-half of a percent of the total sum of squares. Although "sums of squares" do not have the same significance for the survey data as for data obtained by simple random sampling, it is still possible to get an approximation to the relative magnitudes of the different effects.

In table IV of this appendix the entries show these effects for each group, e.g., for the total group (1), the

Table IV. "Between sums of squares" in each group of AID program based on socioeconomic and demographic predictor variables

Group	1	3	2	5	6	8	4	10	9	7	12	13	11
TSS	102.71	59.01	40.18	38.96	29.24	20.76	18.65	17.97	17.10	10.32	9.01	7.40	1.90
Sex	0.864	1.403	0.188	(1)	0.137	(1)	(1)	(1)	(1)	0.108	(1)	(1)	(1)
Race	3.615	1.348	0.620	1.104	0.004	(1)	0.324	(1)	(1)	(1)	(1)	(1)	(1)
Geographic region	1.814	0.977	0.128	0.862	0.046	0.258	0.218	0.191	0.529	0.218	0.255	(1)	(1)
Rate of population change	0.194	0.038	0.013	0.222	0.022	0.021	0.073	0.028	0.133	0.053	0.011	0.011	(1)
Size of place of residence	0.193	0.178	0.012	0.160	0.006	0.067	0.033	0.100	0.691	0.123	0.084	0.000	(1)
SMSA or non-SMSA	0.028	0.028	0.007	0.109	0.015	0.013	0.040	0.000	0.460	0.013	0.030	0.011	(1)
Number of youths in household	1.329	1.058	0.195	0.982	0.111	0.506	0.512	0.444	0.262	0.051	0.203	0.150	(1)
Age	0.381	0.375	0.110	0.328	0.060	0.202	0.125	0.204	0.270	0.241	0.367	0.011	(1)
Education of first-listed parent	2.475	0.770	0.476	1.045	0.306	0.896	0.083	0.221	0.170	0.079	0.107	0.083	(1)
Family income	3.638	0.360	0.584	0.227	0.244	0.036	0.188	0.010	0.103	0.324	0.000	0.246	(1)
Number of sample youths	6,755	1,761	4,994	876	4,598	586	885	561	290	396	190	100	25
Rate of illiteracy	4.76	11.7	2.4	16.7	1.9	12.4	6.8	11.0	25.9	9.3	19.4	37.8	46.0

[1]Indicates that the group cannot be split on the basis of this predictor variable.

NOTE: The lines ↘ start in the column of the group which was split and the row of the predictor variable on which the split was based. The arrows indicate the resulting groups.

BSS (between sum of squares) for income is the largest, race and education of first-listed parent being the next two. Another way of expressing this would be that in group (1) about 3.5 percent of the total sum of squares (TSS) is explained by the effect of family income. When a split has been made on the basis of family income into groups (3) and (2), the ordering of the BSS changes from that in (1) and in a different way for each group. In group (3) sex explains about 2 percent of the variation in that group (i.e., out of the TSS of 59.01) whereas race explains a little less. In group (2) race accounts for much more variation than sex. It should also be mentioned that when a dichotomization is performed, the TSS for the group being split is thereby partitioned into the sum of the TSS for the resulting group plus the BSS for the variable on which the split was based, e.g., when group (3) is split on the basis of sex into groups (4) and (5), one can see that TSS(3)—the total sum of squares for group (3)—can be decomposed in the following way:

$$TSS(3) = TSS(5) + TSS(4) + BSS(3, \text{sex}) \quad \text{or}$$

$$59.01 = 38.96 + 18.65 + 1.403$$

In the case under discussion the process split up the data into seven "terminal" groups. Examining the cumulative BSS for the terminal groups as a proportion of the original sum of squares, [i.e., the TSS for group (1)] reveals that only about 8 percent of the total variation is accounted for by the dichotomization. The "within sum of squares" in each group is still very large and accounts for 92 percent of the variation. If the restriction of $F2$ is changed to .003 the process is carried on into 12 terminal groups but the percentage of explained variation increases to about 10 percent only. In other words the rates as found by the Brief Test can only to a small extent be explained in terms of the demographic and socioeconomic variables considered in the survey.

APPENDIX III

DEFINITIONS OF CERTAIN TERMS USED IN THIS REPORT

Age.—A youth's age as of his last birthday on the date of first interview was used as the basis for deciding whether or not he was to be included in the sample. However, the age recorded for each youth was his age as of his last birthday on the date of examination. Age was confirmed by comparison with the date of birth given on the youth's birth certificate. Since the examination usually took place 2-4 weeks after the interview, some of those who were 17 years old at the time of interview had become 18 years old by the time they were examined. There were 58 such cases. In the adjustment and weighting procedures these cases were included in the 17-year-old group.

Race.—The race classification recorded by observation was confirmed whenever possible by comparison with the race classification on the youth's birth certificate. Race was recorded as white, Negro, or all other. "All other" included American Indians, Chinese, Japanese, and all races other than white or Negro.

Parent.—A parent was the natural parent or, in the case of adoption, the legal parent of the youth.

Guardian.—A guardian was the person responsible for the care and supervision of the youth. She (or he) did not have to be the legal guardian to be considered the guardian in this survey. A guardianship could exist only when neither parent resided in the sample household.

Family income.—The income recorded was the total income received during the past 12 months by the head of the household and all other household members related to the head. This income was the gross cash income (excluding pay in kind) except in the case of a family with its own farm or business, when net income was recorded. Also included in the family income figure were allotments and other money received by the family from a member of the Armed Forces.

Education of parent or guardian.—The highest grade that had been attended in school was recorded. Also recorded was whether that grade had been completed. The only grades counted were those which had been completed in a regular public or private school where persons were given formal education, whether during the day or at night and whether attendance was full time or part time. A "regular" school is one which advances a person toward an elementary or high school diploma or a college, university, or professional school degree. Education or training received in vocational, trade, or business schools outside the regular school system was not counted in determining the highest grade of school completed.

Geographic region.—The United States was stratified into four broad geographic regions of approximately equal population. These regions, which deviate somewhat from those used by the U.S. Bureau of the Census, are as follows:

Region	*States Included*
Northeast	Maine, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, and Pennsylvania
South	Delaware, Maryland, District of Columbia, West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, and Arkansas
Midwest	Ohio, Illinois, Indiana, Michigan, Wisconsin, Minnesota, Iowa, and Missouri
West	Washington, Oregon, California, Nevada, New Mexico, Arizona, Texas, Oklahoma, Kansas, Nebraska, North Dakota, South Dakota, Idaho, Utah, Colorado, Montana, Wyoming, Alaska, and Hawaii

Urban and rural areas.—The classification of urban and rural areas was that used in the 1960 Census. According to the 1960 definition, those areas considered urban were (a) places of 2,500 inhabitants or more incorporated as cities, boroughs, villages, and towns (except towns in New England, New York, and Wisconsin), (b) the densely settled urban fringe, whether incorporated or unincorporated, of urbanized areas, (c) towns in New England and townships in New Jersey and Pennsylvania which contained no incorporated municipalities as subdivisions and had either 2,500 inhabitants or more, or a population of 2,500 to 25,000 and a density of 1,500 persons or more per square mile, (d) counties in States other than the New England States, New Jersey, and Pennsylvania that had no incorporated municipalities within their boundaries and had a density of 1,500 persons or more per square mile, and (e) unincorporated places of 2,500 inhabitants or more which were not included in any urban fringe. The remaining population was classified as rural.

Rate of population change.—The rate of population change measured the increase or decrease of population between 1950 and 1960. The primary sampling unit in which the youth resided was classified as belonging to one of four groups designated "population loss," "below average gain," "average gain," or "above average gain." However, these classes were defined differently for each geographic region. Details are given in *Vital and Health Statistics*, Series 2, No. 43.[5]

★ U. S. GOVERNMENT PRINTING OFFICE : 1973 543-070/56

VITAL AND HEALTH STATISTICS PUBLICATION SERIES

Originally Public Health Service Publication No. 1000

Series 1. Programs and collection procedures.—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions, data collection methods used, definitions, and other material necessary for understanding the data.

Series 2. Data evaluation and methods research.—Studies of new statistical methodology including: experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, contributions to statistical theory.

Series 3. Analytical studies.—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. Documents and committee reports.—Final reports of major committees concerned with vital and health statistics, and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. Data from the Health Interview Survey.—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, based on data collected in a continuing national household interview survey.

Series 11. Data from the Health Examination Survey.—Data from direct examination, testing, and measurement of national samples of the civilian, noninstitutional population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics; and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. Data from the Institutional Population Surveys.—Statistics relating to the health characteristics of persons in institutions, and their medical, nursing, and personal care received, based on national samples of establishments providing these services and samples of the residents or patients.

Series 13. Data from the Hospital Discharge Survey.—Statistics relating to discharged patients in short-stay hospitals, based on a sample of patient records in a national sample of hospitals.

Series 14. Data on health resources: manpower and facilities.—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. Data on mortality.—Various statistics on mortality other than as included in regular annual or monthly reports—special analyses by cause of death, age, and other demographic variables, also geographic and time series analyses.

Series 21. Data on natality, marriage, and divorce.—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports—special analyses by demographic variables, also geographic and time series analyses, studies of fertility.

Series 22. Data from the National Natality and Mortality Surveys.—Statistics on characteristics of births and deaths not available from the vital records, based on sample surveys stemming from these records, including such topics as mortality by socioeconomic class, hospital experience in the last year of life, medical care during pregnancy, health insurance coverage, etc.

For a list of titles of reports published in these series, write to:

Office of Information
National Center for Health Statistics
Public Health Service, HRA
Rockville, Md. 20852

DHEW Publication No. (HRA) 74-1613
Series 11 - No. 131

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE
Health Resources Administration
5600 Fishers Lane
Rockville, Maryland 20852

OFFICIAL BUSINESS
Penalty for Private Use $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF HEW

HEW 390

U.S.MAIL

THIRD CLASS
BLK. RT.